A Graph Cut Approach to Artery/Vein Classification in Ultra-Widefield Scanning Laser Ophthalmoscopy

Pellegrini, Enrico; Robertson, Gavin; MacGillivray, Tom; van Hemert, Jano; Houston, John; Trucco, Emanuele

Published in:
IEEE Transactions on Medical Imaging

DOI:
10.1109/TMI.2017.2762963

Publication date:
2018

Document Version
Peer reviewed version

A Graph Cut Approach to Artery/Vein Classification in Ultra-Widefield Scanning Laser Ophthalmoscopy

Enrico Pellegrini, Gavin Robertson, Tom MacGillivray, Jano van Hemert, Graeme Houston and Emanuele Trucco

***Abstract*—The classification of blood vessels into arterioles and venules is a fundamental step in the automatic investigation of retinal biomarkers for systemic diseases. In this paper we present a novel technique for vessel classification on ultra-wide-field-of-view images of the retinal fundus acquired with a scanning laser ophthalmoscope. To our best knowledge, this is the first time that a fully automated artery/vein classification technique for this type of retinal imaging with no manual intervention has been presented. The proposed method exploits hand-crafted features based on local vessel intensity and vascular morphology to formulate a graph representation from which a globally optimal separation between the arterial and venular networks is computed by graph cut approach. The technique was tested on three different datasets (one publicly available and two local) and achieved an average classification accuracy of 0.883 in the largest dataset.**

***Index Terms*—Retina, scanning laser ophthalmoscope, ultra-widefield, artery-vein classification, blood vessel, graph cut.**

I. Introduction

IMPROVEMENTS in computerized imaging modalities have made available to clinicians increasing volumes of digital high-resolution images of the retina for the investigation of eye and systemic diseases. In the retina, a rich portion of the microvasculature of the human body can be observed non-invasively. Hence, retinal microvascular features and their changes have been investigated as candidate biomarkers in association with different types of systemic conditions such as cardiovascular disease, dementia and diabetes [1]–[5].

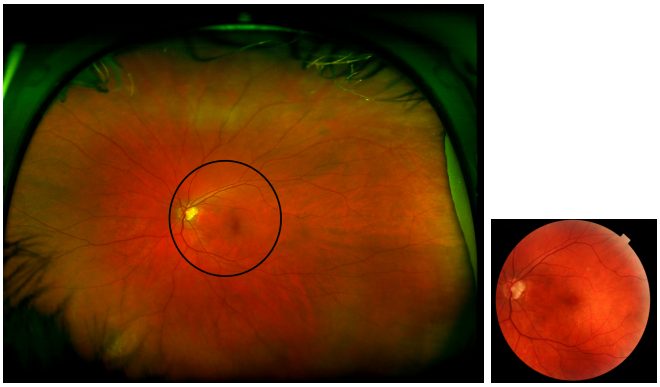
Studies of retinal biomarkers for systemic diseases have traditionally been restricted to a small circular annulus centred on the optic disc (OD) observed in fundus camera images [6]–[9]. Acquiring single images of extended retinal regions (beyond ~45°) was possible only via repeated acquisitions and montages, the prime example being the Early Treatment Diabetic Retinopathy Study (ETDRS) group protocol for diabetic retinopathy [10]. The introduction of the scanning laser ophthalmoscope (SLO) [11] made it possible to capture in one shot (Fig. 1) an image with an ultra-widefield of view (UWFoV) of 180-200° [12], reducing patient discomfort while guaranteeing good image resolution.

Pellegrini and Trucco are with the VAMPIRE Project, School of Science and Engineering (Computing), University of Dundee, Dundee, DD14HN, UK. E-mail: e.z.pellegrini, e.trucco@dundee.ac.uk.

Robertson and MacGillivray are with the VAMPIRE Project, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, EH16 4TJ, UK. E-mail: t.j.macgillivray, gavin.robertson@ed.ac.uk.

van Hemert is with Optos plc, Dunfermline, KY11 8GR, UK. E-mail: jvanhemert@optos.com

Houston is with Ninewells Hospital & Medical School, University of Dundee, Dundee, DD1 9SY, UK. E-mail: J.G.Houston@dundee.ac.uk



(a) (b)

Fig. 1. **UWFoV SLO vs Fundus camera.** (a) Example of UWFoV image, (b) fundus camera image of the same eye. The black circle in (a) represents approximately the FoV visible in (b).

An automated system providing measurements from UWFoV images for biomarker research must segment and classify small retinal blood vessels reliably as arterioles or venules. This operation is very time-consuming if performed manually, therefore it is very important to develop software tools capable of carrying it out effectively.

We propose a novel automated technique for artery / vein (AV) classification of the retinal blood vessels visible in UWFoV SLO images. We first extract hand-crafted features along the vessel centrelines and train a local classifier from which AV labels can be derived first at pixel level and then at segment level. The classification accuracy is then significantly improved by computing a representation of the network of the blood vessel segments as an undirected graph and partitioning the network ensuring the consistent propagation of the labels at a global level. This is achieved by making use of a novel formulation of soft constraints and edge costs for a graph cut approach, aimed at disambiguating branching and crossing point configurations throughout the entire vasculature network.

The strengths of the proposed method are:

- Accepting as input imperfect binary vessel maps computed in a fully automatic way from real UWFoV images, with no manual adjustment. To our best knowledge, the only other comparable system reported using UWFoV images [13] on the subject was validated on manually specified centreline maps only.
- Globally optimal AV labelling is enforced to minimize label inconsistencies along vessel paths.
- No fixed rules, based on local features of the vascular morphometry, are pre-determined to disentangle crossing

and bifurcation points. Instead, the parameters to carry out this step were optimized on a separate set of images.
- Information regarding the branching and crossing points morphology and the vessel local appearance are combined and exploited simultaneously for the graph edge classification stage, instead of being set according to fixed *a priori* rules as in similar studies [14].

To our best knowledge, this is the first time that an AV classification algorithm is applied to obtain a global labelling in UWFoV SLO images starting from an automatically generated segmentation of blood vessels.

II. Related work

A. Local classification methods

Several automated techniques have been proposed for AV classification in conventional fundus camera images. Individual steps differ considerably from one method to another but four main phases can be found in the majority of the approaches. First, the vessels are segmented, either with an automatic or a manual tool. Second, the resulting binary map is pruned and skeletonized in order to extract the centrelines of the blood vessels. Third, a set of local features is computed at each centreline pixel. Fourth, the feature vectors are used to classify the pixels by either supervised or unsupervised approaches and the resulting labels are combined to obtain labels for whole vessels segments.

A well-known, early study by Grisan and Ruggeri [15] addressed the task in a small circular region around the OD. The authors obtained 0.93 classification accuracy for the main vessels by utilizing a *divide et impera* strategy and selecting only two colour features. A similar approach based on clustering of features was proposed by Vazquez *et al.* [16] and then extended, with a focus on features representative of the cross-sectional profile of the blood vessels, by Saez *et al.* [17]. Colour features were also used in other unsupervised [18] and supervised [19] algorithms for AV classification. Methods by Niemeijer *et al.* [20], Zamperini *et al.* [21] and Mirsharif *et al.* [22] explored larger sets of features extracted from single pixels, regions of interest (ROI's) centred on the vessel pixel in question and from vessel segments. Feature selection was then performed and a number of different classifiers were tested to assess which one gave the best performance.

We built on the paper by Zamperini *et al.* [21] to devise a large set of hand-crafted features, discriminative at pixel-level. In the framework proposed in [21], intensity values and their first- and second-degree moments were computed in the channels of different colour spaces (RBG and HSV). The values were extracted twice: first in a ROI centred on the pixel in question and with a diameter equal to the width of the vessel at that point, then in a ROI twice as large as the previous one. Also, information regarding the location of the pixel seemed to improve classification accuracy. Finally, feature selection was carried out and several types of classifiers were trained: the one achieving the best results was the linear Bayesian classifier (LBC) [23].

B. Graph-based methods

More recently, the properties of the entire vasculature have been taken into account to gain additional discriminative information. The correct representation of vessels in the segmented binary maps and the disentangling of ambiguous branching and crossing points was investigated by Al-Diri *et al.* [24] and by Huang *et al.* [25]. In these two papers, though, the results were not evaluated in terms of AV classification. The first representation of the vasculature as a graph specifically designed for AV classification purposes was the semi-automated one proposed by Rothaus *et al.* [26]. In this case, two paired graphs were devised and the results were evaluated in terms of labelling conflicts that were solved by the algorithm.

More recently a growing number of studies, among which [14], [27]–[31] and [13], have proposed graph representations to tackle the AV classification problem. Estrada *et al.* [13] reported the first AV classification in UWFoV SLO images. This technique is an extension of a previous paper from the same authors [32], based on a heuristic optimization algorithm to explore the space of anatomically plausible vascular networks. A global likelihood model based on vessel tree features was devised for each one of these configurations. In these approaches, the structure of the vasculature is represented as a graph and exploited in order to ensure that the proposed AV labels are consistent across the vascular trees and over the entire FoV of the image.

Joshi *et al.* [14] started with the segmentation of blood vessels and the creation of a refined centreline map where vessel segments were connected at branching and crossing point. They modelled the segments as nodes of a graph and assigned to the graph edges a set of weights based on local features. A partition of the graph was then obtained using the Dijkstra algorithm. Labels were assigned to graph components using fuzzy C-means clustering of a set of 4-D feature vectors extracted at pixel-level. The vector components were four colour features (mean and standard deviation of green channel and hue) which proved to be very discriminative in the fundus camera images considered. It is interesting to observe that excluding pixels with very uncertain AV labels was considered infeasible in UWFoV SLO image as it would have led to the elimination of the majority of the data.

Finally the technique by Dashtbozorg *et al.* [28] was relevant for this work for the preprocessing steps put in place. The rest of the work was similar to the one by Joshi *et al.* in terms of modelling AV classification as a graph partitioning problem. The most notable differences were that the actual separation was obtained by following a set of rules, defined empirically, to solve all crossing and branching points in sequence and that the local AV labels were devised by creating a larger feature vector and using a linear discriminant classifier. Preprocessing consisted of a set of fixed rules used to correct segmentation errors visible in the centreline map. These empirical rules were based on threshold values obtained by combining local features of vessel segments. It is worth noting that the work by Dashtbozorg *et al.* was one of the few graph-based algorithms proposed in the literature that was validated and able to perform reliably on centreline maps not obtained, or refined,

TABLE I
DATASETS USED TO ASSESS THE PERFORMANCE OF THE AV CLASSIFICATION ALGORITHM.

Name	Number of images	Vessel segments	Centreline pixels	Number of observers	Inter-observer agreement	Location of segments	Image resolution*	Stereographic projection
ZONE C	40	11.6 (1.2)	237.3 (31.2)	2	100 %	Zone C	Full	No
WIDE	30	48.0 (7.3)	856.9 (74.9)	2	98 %	Entire FoV	Half	No
TASCFORCE	470	113.5 (26.3)	1572.3 (112.5)	1	n. a.	Entire FoV	Full	Yes

* For a non-projected UWFoV SLO image, "Full" resolution represents to an image of size 3072×3900 pixels covering a FoV of approximately $180°$. The respective size of a stereographically projected UWFoV SLO image covering the same FoV is 4000×4000 pixels. Vessel segments and centreline pixels are reported as mean (standard deviation) per image.

manually unlike in [13], [32].

III. MATERIALS

The proposed AV classification algorithm was tested on three sets of UWFoV SLO images differing in terms of image resolution, number of observers, FoV taken into consideration and whether the images underwent a stereographic projection [33] or not. A summary of the dataset characteristics is presented in Table I.

The ZONE C dataset consists of 40 images captured from healthy volunteers with a very low incidence of retinal pathologies in 2013 at Ninewells Hospital (Dundee, UK) with a P200C AF ophthalmoscope [11] (Optos plc, Dunfermline, UK). The protocol involved non-mydriatic acquisition of 2 good-quality images, left and right, per subject. This set was manually annotated by two trained observers who labelled as arteries or veins up to 12 vessel segments per image. The vessel selection was limited to zone C, defined as the annulus between 0.5 and 2.5 optic disc diameters (ODD's) from the OD margin [5].

The 470 images (470 subjects, one image per subject) of the TASCFORCE [34] set used in the experiments were captured in the same setting as above but the two are not overlapping. In the TASCFORCE set, the vessel segments that were manually labelled, by only one observer because of time constraints, were those included in the vessel binary map produced by our automated technique [35].

The WIDE dataset was made publicly available by Estrada *et al.* [13] and consists of 30 UWFoV SLO images that were previously cropped to a smaller FoV and scaled down by a factor of two, resulting in a lower resolution with respect to the other sets used in the experiments. It is very important to note that the WIDE dataset comprises only maps of vessel centreline points that are labelled as arterioles or venules. To the best of our knowledge, the full manual segmentation of the blood vessels is not available, hence no information regarding the local vessel width is provided. In this work, the images from the WIDE dataset were first scaled back to their original size by means of a bicubic interpolation and then processed. The reasoning behind this choice was that the entire pipeline, including the preprocessing steps, was optimized to work at the native image size of the UWFOV SLO device. The only vessels, among those automatically segmented, considered in the experiments were those that had been labelled by the two observers as belonging to the same class.

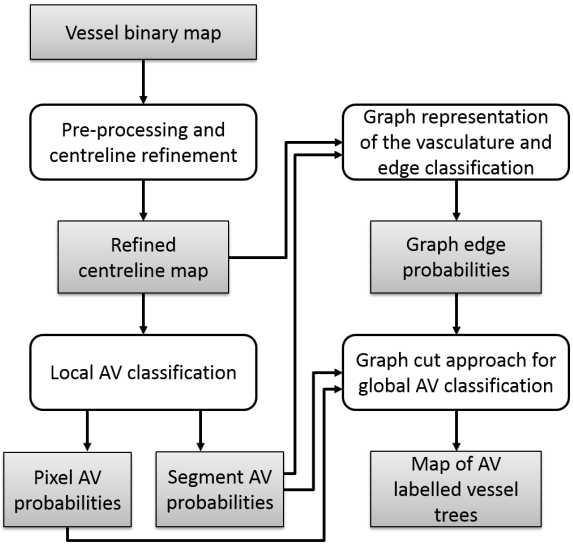


Fig. 2. **Pipeline of the proposed method.** White boxes indicate the algorithm's modules, grey boxes indicate input and output of the modules.

IV. METHODS

The pipeline of the proposed algorithm can be conceptually divided into four interconnected modules (Fig. 2). Each module is detailed in one of the following subsections.

A. Pre-processing and centreline refinement

A binary map of the blood vessels was obtained for each image by making use of an automated segmentation technique [35]. A circular area, centred on the OD and with a diameter equal to 0.75 ODD's, was masked out of the images as not relevant for the purposes of this study. The map was also thinned [36] in order to obtain the vessel centrelines. This operation left foreground pixels with at most 4 neighbours in a 3×3 neighbourhood. According to the number of neighbours, foreground pixels were labelled as either noise, end points, segment points or meeting points (respectively, pixels with 0, 1, 2, more than 2 neighbours). Vessel segments were then defined as all the connected components left in the centreline map after the removal of noise, end and meeting points.

Subsequently, the centreline map underwent an automated refinement aimed at correcting misinterpretations of the vascular structure introduced by the automated segmentation technique or by the thinning operation. Similar to [28], three types of refinement were introduced: the creation of a missing vessel

TABLE II
FEATURES FOR LOCAL AV CLASSIFICATION.

Feature number	Description
1-6	Centreline pixel intensity in Red, Green, Hue, Value, Greyscale and Gradient maps.
7-30	Mean, standard deviation, minimum and maximum values in the sROI in Red, Green, Hue, Value, Greyscale and Gradient maps.
31-54	Same as (7-30) but calculated in the lROI.
55-64	Centreline pixel intensity in five Gaussian-blurred versions ($\sigma = [2\ 4\ 8\ 12\ 16]$) of Red and Green maps.
65-69	Polar coordinates* of the centreline pixel.
70	Vessel width at the centreline pixel location.
71-80	Vessel cross-sectional profile† in the Red map.
81-90	Vessel cross-sectional profile† in the Green map.

* Three reference systems: centred on the OD with the vertical axis of the image, centred on the OD with the perpendicular to the direction OD-fovea as the axis, centred on the image centre with the vertical axis of the image. The radial coordinates were normalized according to the individual ODD.
† The profile length was determined as 140% of the estimated vessel width, the intensity was approximated by a cubic spline interpolation and then sampled at 10 equally spaced locations.

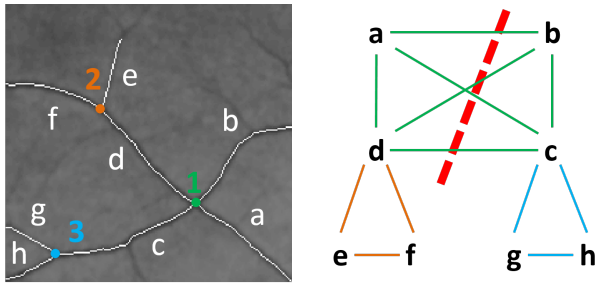
segment, the deletion of a false segment and the correction of a split crossing point. The set of rules and threshold values used to identify the locations needing refinement were not pre-determined but devised as functions of a number of parameters obtained from a training set in order to minimize the occurrence of "wrong corrections" (e.g., creating a new segment where there should not be one). This operation was carried out by exhaustive search on a separated small set of 20 UWFoV SLO images and validated by visual inspection with the aim of achieving the lowest possible false discovery rate.

The use of a vessel map obtained automatically, always less accurate than a manual map, and its refinement are crucial to highlight that the proposed method is resilient to the errors introduced by automatic vessel segmentation. This achieves for the first time for UWFoV SLO images a pipeline for AV classification requiring no manual interaction throughout the process.

B. Local AV classification

After obtaining a refined centreline map, the vessel segments were sub-sampled at equally-spaced intervals: 1 sample every 10 points along the centreline. Pixels that were edited during the previous stage of automatic centreline refinement were excluded from the pool of candidate samples. The final number of pixels taken into consideration for further analysis in each dataset is reported in Table I.

For each of the selected pixels, a hand-crafted 90-D feature vector (Table II) was created. Twenty three of the features were calculated twice: once in a small circular region of interest (sROI) centred on the point, with a diameter equal to the vessel width and once in a larger, concentric region (lROI) with a diameter twice as large [21]. With this choice, the sROI and its features were strictly representative of the vessel pixels



(a) (b)

Fig. 3. **Graph representation of the vasculature.** (a) Green channel with vessel centrelines highlighted in white and where 8 segments and 3 meeting points are visible, (b) Respective graph representation. The red dashed line covers the edges belonging to the optimal cut C.

while the lROI gave information of both the vessel pixels and its surrounding background. Location information, estimated local width [37] and cross-sectional profile intensities were finally added to complete the feature vector.

The local AV classification of the centreline pixels was carried out independently for each set of images, in each case performing a leave-one-image-out cross-validation. The training sets were sub-sampled to ensure class balance between arteriolar and venular pixels and the features rescaled into the $[0,\ 1]$ range. A LBC [23] and the entire 90-D feature vector were used so that at the end of the process, every pixel p_i was given a probability $P_a(p_i)$, also referred to as A-probability, of belonging to the arteriole class. These values were thresholded at 0.5 to obtain the individual AV labels $L(p_i)$ that were taken in consideration in the evaluation of the classification accuracy on pixels. Finally, the A-probabilities of the pixels belonging to each vessel segments were averaged to obtain segment-level A-probabilities. The latter were thresholded as well and the resulting values were considered in the evaluation of classification accuracy on segments.

Preliminary tests were run on the ZONE C dataset to investigate classification accuracy values achievable by combining different feature selection techniques (1-way ANOVA, elastic net [38], LASSO [39]) and classifiers (ordinary and regularized least squares, linear regression, linear and RBF SVM's). The results of these experiments are reported in Appendix A. Ultimately, the LBC with no feature selection proved to be the most accurate combination (Table IX) and was thus used, from this point forward, for the rest of the experiments.

C. Graph representation of the vasculature and n-edge classification

Each connected component of the refined centreline map was modelled as an undirected graph where the vessel segments were represented by nodes. Segments sharing a meeting point were connected by an edge in the graph model (Fig. 3). The graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ was the set of nodes (or vertices) and $\mathcal{E}$ the set of edges, was defined accordingly. The final goal of the proposed method was to identify the cut $C \subset \mathcal{E}$ (Fig. 3b) such that all the nodes representing respectively venules and arterioles became separated in the

induced graph $G(C) = (\mathcal{V}, \mathcal{E} \setminus C)$. By deleting all the edges belonging to the cut, obtained by making use of a max-flow algorithm [40], an optimal AV classification, consistent at global level, is achieved.

Two additional nodes, the source s and the sink t, were added to the graph and fully connected to all the other nodes. Following the notation in [41], edges between neighbouring nodes of the original set $\mathcal{V}$ were defined as *n-edges*. The edges connecting all the nodes in $\mathcal{V}$ and the terminals s and t were instead defined as *t-edges*. A weight was associated to each edge as an indication of the cost of cutting it. A high cost was imposed to *n-edges* edges joining vessel segments of the same AV class and a low cost to those between an arteriole and a venule. Also, high and low costs, respectively, were assigned to *t-edges* between nodes representing venule segments (*vice versa* for arteriole segments) and the terminals s and t. The cost of all the edges was determined by the output of the previous step of local AV classification and by the output a set of *n-edges* classifiers aimed at discriminating between Class 0 *n-edges* (e.g., links between an arteriole and a venule) and Class 1 *n-edges* (i.e., links between segments belonging to the same AV class). This second stage of classification is described in the following subsection.

To determine the class of a *n-edge*, it was not sufficient to focus on the information regarding the two vessel segments that formed it. It was also necessary to gather information representative of the configuration of the meeting point from which the edge was originated. The final feature vector of each *n-edge*, therefore, was the concatenation of a set of *segment features* and a set of *meeting point features*. This constitutes a novelty aspect of the proposed method with respect to previous techniques where, to make sense of the morphology of the vasculature, only local geometrical features were considered [28] or a hierarchy of features was determined *a priori* [14] by an observer.

The meeting points were divided in subgroups according to their degree, defined as the number of vessel segments attached to them. A different LBC, based on different hand-crafted features, was trained to classify edges from each one of these subgroups. Across the entire WIDE and TASCFORCE image sets, only four configurations were detected as the maximum degree of a meeting point was found to be 5. The feature vectors for the edges from configuration of degree 2, 3 and 4 are reported in Table III. Since only five occurrences of 5-degree configurations were detected across all the dataset, not enough data was available to train a classifier for edges originating from this type of configuration. In these few cases, arbitrarily set values of probability were assigned to the edges: the edge between the two segment separated by the angle closest to 180° was associated with a probability equal to 0.8, while 0.5 was the probability assigned to the remaining edges.

TABLE III
FEATURE VECTORS FOR *n-edge* CLASSIFICATION GROUPED BY MEETING POINT'S DEGREE.

Degree	Type of features	Feature number	Description
2	Segment features	1	Difference of segments' AV probabilities.
		2	Smallest angle between segments.
		3	Difference of segments' average Euclidean distances from the OD.
3	Segment features	1, 2	A-probabilities of the two segments.
		3-5	Difference of widths and intensities on the red and the green channel between the two segments.
		6	Smallest angle between the two segments.
		7, 8	Segments' Euclidean distances from the OD.
	Meeting point features	9, 10	Angles between the two segments in question and the main segment*.
		11	Angles between the third segment and the main segment.
		12	Number of end segments† attached to the meeting point.
		13	A-probability of the third segment.
4	Segment features	1-8	Same as those for Degree = 3.
	Meeting point features	9, 10	Same as those for Degree = 3.
		11, 12	Angles between the other two segments and the main segment.
		13	Ranking of the angles between the two segments in question with respect to the rest of angles between segment couples.
		14, 15	A-probability of the remaining segments.
		16	Number of segments with a Euclidean distance from the OD larger than the one of the meeting point.

* The main segment of the configuration was defined as the one with the shortest distance from the OD.
ˆ The third segment was defined as the only segment connected to the meeting point not touch by the edge in question.
† An end segment was defined as a segments connected to one end point.

D. Graph cut approach for global AV classification

Once the undirected graph representation was obtained for each connected component of the retinal vasculature and the two classification tasks (AV and *n-edge*) were completed, the two vessel segments with the highest and the lowest A-probability were identified, respectively, as seeds s_1 and s_2. For each node $v_i \in \mathcal{V}$, the cost R_{v_i} of its *t-edges* was calculated according to Table IV, using as starting point the sets of probabilities previously discussed.

In particular:

$$C_i = (1 + \sum_{v \in \mathcal{N}_{s_i}} P_e(s_i, v)),\ i \in [1, 2], \tag{1}$$

where $P_e(s_i, v)$ was the probability of the *n-edge* between nodes s_i and v of being a Class 1 *n-edge* and $\mathcal{N}_v$ (neighbour-

TABLE IV
COST OF *t-edges* IN THE GRAPH MODEL.

Node	Edge with s	Edge with t
s_1	$\lambda_1 C_1$	$(1-\lambda_1)C_1$
s_2	$(1-\lambda_1)C_2$	$\lambda_1 C_2$
$v \in \mathcal{V} : v \neq s_1, s_2$	$\lambda_2 P_a(v)$	$\lambda_2(1-P_a(v))$

hood) was the set of nodes $v \in \mathcal{V}$ that shared an edge with s_i. Finally, $P_a(v_i)$ was the A-probability of v_i.

All the R_{v_i} values were rescaled preserving the resolution in order to make them positive integers. The coefficients λ_1 and λ_2 were the two model parameters to be optimized.

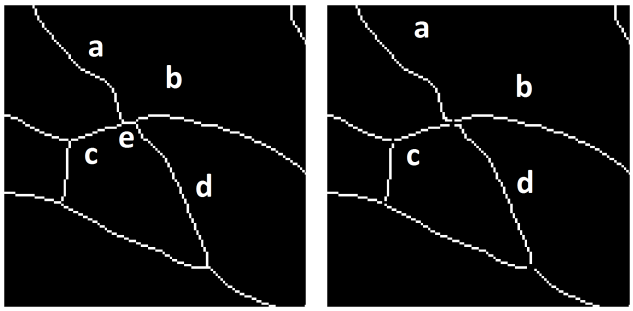
A novel aspect of the proposed technique is that, differently from the approach in [41], it does not rely on a pre-defined set of seed points that result in enforcing of hard constraints on the graph formulation. The seeds used in the proposed method were not selected manually but were rather obtained from the first stage (i.e., the local AV classification) of the method itself. Since they were the result of a classification technique with a certain level of accuracy, the constraints forcing the seeds to belong to a specific AV class were relaxed according to the coefficient $\lambda_1 \in [0.5, 1]$. This coefficient was introduced to represent the confidence in the selection of the seed nodes, i.e. the ability of the local AV classification technique to assign very high or very low A-probabilities to the more representative segments of the two classes. For example, $\lambda_1 = 1$ would indicate that the local AV classification is considered always correct in the assignment of the highest and the lowest A-probability in a set of vessel segments. In that case, the cost of severing the edge (s_1, s) connecting s_1 to the terminal s would be higher than the total cost of severing the edge (s_1, t) and all the *n-edges* between s_1 and its neighbours combined. The opposite would happen for seed s_2.

The second coefficient $\lambda_2 > 0$ specified the relative importance of the A-probabilities over the edge probabilities. It could therefore be considered an indication of the confidence on the accuracy of the AV and the edge classifiers.

The cost $B(v_i, v_j)$ of the *n-edges* between v_i and each other node v_j in its neighbourhood was calculated according to the following rule:

$$B(v_i, v_j) = \begin{cases} 1 & \text{for } \textit{true edges} \\ 0 & \text{for } \textit{false edges} \\ P_e(v_i, v_j) & \text{for all other edges} \end{cases}, \quad (2)$$

where *true* and *false* edges were defined as those affected, and thus respectively deemed correct or not, by the automatic corrections of the centreline map during the pre-processing stages. For example, consider the case of a crossing point (4-degree) between an arteriole and a venule that was first represented in the centreline map as two different 3-degree meeting points (Fig. 4a). If this wrong configuration was recognized during pre-processing and automatically corrected (Fig. 4b), the relevant 6 *n-edges* in the respective graph representation would be set to four *false edges*, associated with a cost equal to 0, and two *true edges*, associated with a cost equal to 1. It is worth noting this aspect of the proposed method since it highlights how deep are the repercussions of the automated pre-processing operations on the graph formulation.



(a) (b)

Fig. 4. **Centreline map refinement and definition of *true* and *false n-edges*.** (a) Example of 4-degree meeting point wrongly represented by two 3-degree meeting points connected by segment *e*. (b) Centreline map after automated refinement: segment *e* is deleted and the two meeting points are merged into a single one. In the respective graph representation, the edges between the couples *a-d* and *b-c* are defined as *true edges* and the edges *a-b*, *a-c*, *b-d* and *c-d* as *false edges*.

All the costs $B(v_i, v_j)$ were then rescaled consistently with the costs R_{v_i}.

For the final formulation, let $L = [L_1, \ldots, L_v, \ldots, L_{|\mathcal{V}|}]$ be a binary vector where each component L_v specified the AV label assign to each vessel segment $v \in \mathcal{V}$. The vector L defined a global AV classification of the entire vasculature. The cost function to be minimized to solve the problem was:

$$E(L) = R(L) + B(L), \quad (3)$$

where

$$R(L) = \sum_{v_i \in \mathcal{V}} R_{v_i}(L_{v_i}), \quad (4a)$$

$$B(L) = \sum_{\{v_i, v_j\} \in \mathcal{E}} B_{v_i, v_j} \; \delta(L_{v_i}, L_{v_j}), \quad (4b)$$

$$\delta(L_{v_i}, L_{v_j}) = \begin{cases} 1, & \text{if } L_{v_i} \neq L_{v_j} \\ 0, & \text{otherwise} \end{cases}. \quad (4c)$$

In the proposed formulation, the coefficients λ_1 and λ_2 did not explicitly appear in (3) but they were embedded in the term $R(L)$ according to Table IV. Their optimization was carried out through an exhaustive search to achieved the highest average value of classification accuracy on pixels at Global level (i.e., the third step of the proposed method) across th entire set of test images in a leave-one-image-out cross-validation setup.

The graph cut algorithm was finally run on all the graph representations in order to divide each one of them into two partitions. The A-probability of each partition was calculated as the average of the A-probabilities $P_a(p_i)$ of the centreline pixels belonging to it. The partition with the highest A-probability was considered as the arterial network while the other partition was labelled as the venular network. A proof of the robustness of the proposed formulation was given by the fact that across all three datasets, no graph after the cut resulted in both partitions having global A-probability values

TABLE V
LOCAL AV CLASSIFICATION RESULTS.

	First step (Pixel-level)				Second step (Segment-level)				
Dataset	Zone C	Central region	Periphery	Total	Zone C	Central region	Periphery	Total	Segment accuracy
ZONE C	0.835 (0.016)	n.a.	n.a.	0.835 (0.016)	0.862 (0.018)	n.a.	n.a.	0.862 (0.018)	0.865 (0.018)
WIDE	0.762 (0.036)	0.755 (0.029)	0.723 (0.035)	0.753 (0.032)	0.812 (0.032)	0.803 (0.028)	0.762 (0.035)	0.801 (0.030)	0.806 (0.031)
TASCFORCE	0.792 (0.017)	0.787 (0.024)	0.756 (0.031)	0.777 (0.025)	0.843 (0.020)	0.830 (0.022)	0.802 (0.029)	0.822 (0.024)	0.827 (0.027)

Accuracy values at the first two steps of the AV classification framework. All columns, except the last one, represent values of accuracy on pixels. The vessel accuracy was obtained by evaluating how many vessel segments were classified correctly after the majority voting step. This metric can be considered as an accuracy weighted according to the length of the blood vessels. Values are expressed as mean and (standard deviation).

higher or lower than 0.5 at the same time.

Finally, the terminal nodes, s and t, were discarded from the graph representation and each connected component of the two partitions was considered as a single vessel tree. The centreline pixels of each tree were assigned a final global A-probability equal to their average value. This probability was then thresholded at 0.5 to obtain a binary label, which was used for the evaluation of AV classification at global (i.e., vascular tree) level.

V. EXPERIMENTAL RESULTS

The proposed technique was evaluated in terms of accuracy of AV classification at different stages of the pipeline and in terms of improvement in the classification of graph *n-edges*. For further comparison, and to highlight the challenges posed by UWFoV SLO images, two AV classification algorithms developed for fundus camera images [18], [21] were tested on the same three datasets. Results are reported in Table X in Appendix A.

A. AV classification results

For each dataset, the values of accuracy at end of each of the first two steps of the proposed method are shown in Table V: the first step was the single centreline pixel classification, the second was the averaging of A-probability values from centreline pixels belonging to the same vessel segment. At the first step of the proposed pipeline, the best results were achieved on the ZONE C dataset. It is worth noting, though, that since the classifier and feature selection combination was chosen according to tests on this set, it is likely that our technique is overfitted to this particular set. For the WIDE and the TASCFORCE sets, the dependency of the accuracy on the distance of the segments from the OD was assessed as well. This evaluation is important since the appearance of arterial and venular vessel segments is more similar in the periphery of the image. Thus, grouped values for pixels located in zone C, in the central region (annulus between $2.5-5.5$ ODDs centred on the OD) and in the periphery of the image are reported. This procedure was not applicable to the ZONE C dataset given the limited FoV (details in Table I). Finally, accuracy values in classifying entire vessel segments are reported in the last column of the table.

TABLE VI
GLOBAL AV CLASSIFICATION RESULTS.

	Third step (Global-level)				
Dataset	Zone C	Central region	Periphery	Total	Segment accuracy
WIDE	0.867 (0.033)	0.865 (0.033)	0.841 (0.029)	0.862 (0.032)	0.864 (0.032)
TASCFORCE	0.896 (0.024)	0.885 (0.023)	0.872 (0.019)	0.882 (0.021)	0.883 (0.024)

Accuracy values at the final stage of the AV classification. All columns, except the last one, represent values of accuracy on pixels. Values are expressed as mean and (standard deviation).

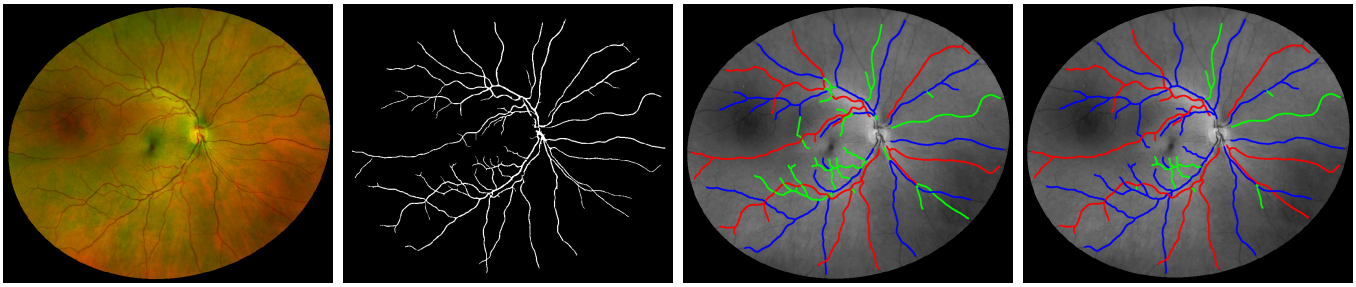
The results of the third step of global AV classification after the graph partitioning for the WIDE and the TASCFORCE sets are reported in Table VI. Consistently with the previous steps of the analysis, the accuracy was evaluated in different regions of the image (zone C, central region and periphery). The same rationale was used to run the experiments reported in Appendix A. Table VI is therefore directly comparable with Table V and Table X.

The global running time for feature extraction and classification of a full size UWFOV SLO image was approximately 20 minutes while the training time for one experiment on the TASCFORCE dataset, by far the largest of the three investigated (Table I), was approximately 10 minutes. This performance was achieved making use of a non-optimized software, coded in MATLAB (ver 2014b), run on a machine with a single i5-3450 CPU @ 3.10 GHz and 8.00 GB of RAM).

B. N-edge classification results

The proposed algorithm was also evaluated in terms of *n-edge* classification accuracy on the WIDE and the TASCFORCE sets separately. In both cases a leave-one-image-out cross-validation was put in place to keep consistency with the AV classification analysis.

The results of the *n-edge* classification task that was carried out before the implementation of the graph cut technique are



(a) (b) (c) (d)

Fig. 5. **Example image at different stages of the process** (a) Original UWFoV SLO image. (b) Binary vessel map. (c) AV labels after the first two steps of AV classification (red = correctly classified arterioles; blue = correctly classified venules; green = wrongly classified vessels). (d) AV labels at the end of the pipeline. The classification accuracy on this image at the second step was equal to 0.783 and after the final step was equal to 0.881. Notice how the graph cut approach is able to improve results in regions characterized by a large number of vessel segments connected to each other but is less powerful in configurations made by few vessel segments or presenting segmentation errors.

reported in the fourth and the fifth columns of Table VII and Table VIII, for the WIDE and the TASCFORCE datasets respectively.

Once the global AV classification was performed, the *n-edges* that belonged to the cut were labelled as Class 0 and the remaining were labelled as Class 1. This final labelling was evaluated in order to show the amount of improvement in edge classification that was due to the graph cut and is reported in the second-last two columns of the tables. Since for each configuration type the two classes were highly unbalanced, the accuracy value for each class is shown. For further comparison with the other algorithms tested, in the last column of the two tables we also report the accuracy in disambiguating 3-degree (i.e., bifurcation points vs wrongly segmented crossing points) and 4-degree meeting points (i.e., crossing points vs wrongly segmented series of two bifurcations). These values are directly comparable with those reported in Table X.

The improvements throughout the three steps of the proposed algorithm are visualized by receiver operating characteristic (ROC) curves for the pixel AV classification on the TASCFORCE set (Fig. 6). The values for the ROCs were obtained by thresholding the A-probabilities at intervals of 0.05 between 0 and 1.

In order to assess the significance of these improvements, two statistical tests were run. A McNemar's [42] test ($\alpha = 0.05$, $p \leq 0.001$) confirmed that the AV classification accuracy achieved at Global-level step of the framework proposed was significantly higher than the accuracy achieved at the segment-level step on the TASCFORCE dataset. This aspect was also confirmed by a Wilcoxon [43] test over the 10-fold cross-validation values of accuracy ($\alpha = 0.05$, $p \leq 0.001$).

VI. Conclusion

In this paper, we have presented a novel multi-stage fully automated technique for the classification of retinal blood vessels in UWFoV SLO images into arteries and veins. To our best knowledge, this is the first time that a fully automatic method is reported to solve this task on UWFoV SLO images, i.e., starting from a vessel map computed automatically from the raw image.

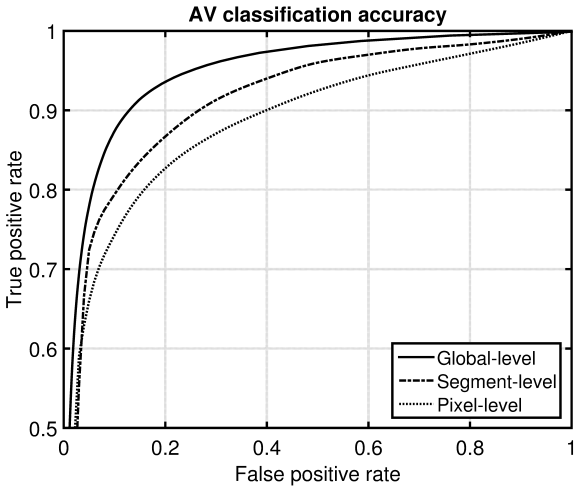


Fig. 6. **AV classification ROCs.** ROC plots of each stage (pixel-level, majority voting, global-level) of pixel AV classification on the TASCFORCE dataset.

The method was tested on three datasets with different characteristics. The results obtained (Table V and Table VI) show consistent performance across the datasets. In every case the second step (segment-level) helped improve the accuracy by correcting the pixels wrongly classified along otherwise correctly labelled vessel segments. In the TASCFORCE and the WIDE sets, a drop in performance in the outer regions of the image was confirmed. This finding was expected since in the periphery vessels may be very thin and colour features less powerful in discriminating between classes. The smaller width contributes in two ways to the deterioration of the method's performance. First, SLO resolution is not enough to enable the retrieval of informative cross-sectional profile features from such small vessels [35]. Second, a higher uncertainty on the width estimation results in a higher uncertainty in the determination of the size of the two regions of interest. This can potentially introduce uncertainty on almost the entire feature vector.

Overall, a significant improvement of performance at each step of the process was recorded for both TASCFORCE and WIDE datasets but there were also relevant differences due

TABLE VII
n-edge CLASSIFICATION ACCURACY FOR WIDE DATASET

Degree	Number of *n-edges*		Edge classification		Final edge labelling		Meeting point disambiguation accuracy
	(Class 0 - 1)		Class 0 accuracy	Class 1 accuracy	Class 0 accuracy	Class 1 accuracy	
2	14	82	0.857	0.914	0.857	0.976	n.a.
3	624	4098	0.776	0.817	0.872	0.899	0.894
4	964	497	0.922	0.931	0.946	0.949	0.946

Accuracy of the classification of *n-edges* in the WIDE dataset at two stages of the framework: after the initial *n-edges* classification task and after the graph cut for the global AV classification. *n-edges* are grouped according to the degree of the meeting point (i.e., number of vessel segments attached to the meeting point) from which they originated. In the last column the power of the algorithm in disambiguating meeting points of a certain degree is reported.

TABLE VIII
n-edge CLASSIFICATION ACCURACY FOR TASCFORCE DATASET

Degree	Number of *n-edges*		Edge classification		Final edge labelling		Meeting point disambiguation accuracy
	(Class 0 - 1)		Class 0 accuracy	Class 1 accuracy	Class 0 accuracy	Class 1 accuracy	
2	256	5095	0.867	0.969	0.867	0.969	n.a.
3	3488	48454	0.827	0.860	0.912	0.930	0.928
4	14776	7778	0.923	0.943	0.964	0.973	0.963

Accuracy of the classification of *n-edges* in the TASCFORCE dataset at two stages of the framework: after the initial *n-edges* classification task and after the graph cut for the global AV classification. *n-edges* are grouped according to the degree of the meeting point (i.e., number of vessel segments attached to the meeting point) from which they originated. In the last column the power of the algorithm in disambiguating meeting points of a certain degree is reported.

to different factors. The WIDE dataset proved to be the most challenging because of its low resolution negatively affecting all the stages of the framework. The vessels were segmented less accurately and the pixel-based features for local AV classification were less discriminative. Both aspects reflected on the performance of edge classification, ultimately resulting in a lower value of global AV classification accuracy.

To our best knowledge, only Estrada *et al.* tried to perform automatic AV classification in UWFoV SLO images but a direct comparison of the proposed method's accuracy and the results (Accuracy = 0.910 for the WIDE dataset) reported in [13] would not be fair considering that the starting point of the two pipelines is fundamentally different: an imperfect vessel map obtained automatically in our case, an optimal map refined by trained observers in the other. The consequences of this difference are visible in Fig. 5d. Where the automatic segmentation fails to detect blood vessels leaving a gap in the vascular network, the AV classification task becomes more challenging. This is especially true when spurious vessel segments are left completely disconnected from the rest of the graph. In such cases, the proposed technique cannot leverage any meaningful information from the vessel morphology and correct the wrong labels obtained at the first step of the pipeline. On the other hand, false positives in the vessel segmentation, are likely to results in crossing point being split into what resembles two "back-to-back" bifurcation points. The use of manually annotated vessel masks would drastically reduce the number of this type of error.
Unfortunately, direct comparison cannot be performed because of the lack of fully manual vessel masks, and the high time cost (estimated 18 hours per image [35]) that would be required to create them, for the WIDE dataset], as previously discussed (Section III).

The results reported here are also lower than those reported in the literature by similar studies involving fundus camera images but also in this case a fair comparison can not be directly made. The two imaging modalities are inherently different and this affects resolution, image contrast, and illuminance. Moreover, one of the strengths of the proposed technique is its power in disambiguating meeting points (Tables VIII, VII and X) and a lower number of these points is visible in the much smaller FoV of a conventional fundus camera image.

Evaluating the edge classification, it can be noticed how the initial accuracy (i.e., before the graph cut consistency enforcement)depends strongly on the degree of the meeting points considered. Edges originated from 2-degree and 4-degree meeting points were classified with a high accuracy from the first stage and experienced only a small improvement after the application of the graph cut technique. The lowest accuracy was instead obtained for 3-degree points. The reason

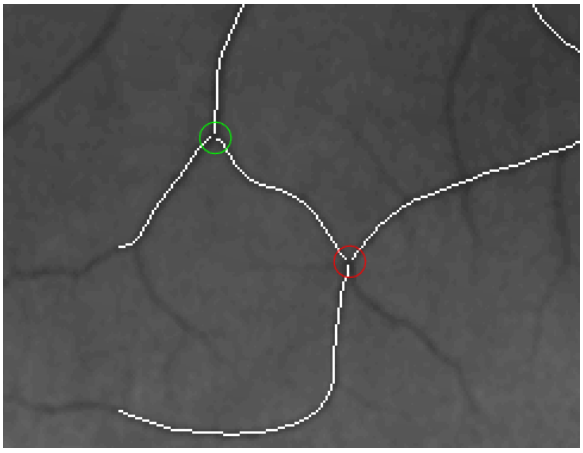


Fig. 7. **Examples of 3-degree meeting points.** Two examples of 3-degree meeting points: actual bifurcation point (circled in green), wrongly segmented 4-degree point resulting in a 3-degree one in the centreline map (circled in red).

has to be attributed to the fact that locally the morphology of a bifurcation in a centreline map is very often indistinguishable from the one of a wrongly segmented crossing point with a missing segment (Fig. 7). The proposed graph cut approach was able to overcome this issue at the final step, achieving a considerable improvement in the labelling of 3-degree Class 0 edges (accuracy from 0.827 to 0.912) reported in Table VIII. In the future, the approach could be improved by exploiting the knowledge of which meeting point configurations are allowed in the centreline map. Disjunctive constraints might be enforced on *n-edges* in the formulation. For example, in a 3-degree configuration, the only possible solutions are that all the three edges are labelled as Class 1 or that only one is. This kind of constraints on graph partitioning and feasible methods to implement them are being investigated in recently reported systems [44], [45].

A final but important observation is that more work is needed to establish what level of accuracy in AV classification is sufficient for supporting a reliable investigation of retinal biomarkers. To our best knowledge, in the main literature this issue has only partially been address in [27], where only the central retinal arterial and venular equivalents and the average tortuosity were considered as outcome measures for validation.

APPENDIX A
ADDITIONAL EXPERIMENTS

The results of the preliminary tests carried out on the ZONE C dataset to investigate different combinations of classifiers and feature selection techniques are reported in Table IX.

The values shown above were obtained by averaging the results of a 3-fold cross-validation. In the same setting, the LBC coupled with no feature selection, achieved a segment-level accuracy equal to 0.847. The abbreviations used in the table are explained below:

- FiltKbets = only the top k features with the highest F-value from a 1-way ANOVA were selected.

TABLE IX
ACCURACY VALUES AT SEGMENT-LEVEL OF DIFFERENT COMBINATIONS OF CLASSIFIERS AND FEATURE SELECTION TECHNIQUES.

	LBC	OLS	RLS	LR	LSVM	RBF
None	0.847	0.819	0.812	0.797	0.8000	0.777
FiltKbest	0.832	0.817	0.821	0.802	0.760	0.779
ENet	0.844	0.840	0.840	0.840	0.842	0.805
LASSO	0.781	0.793	0.794	0.792	0.791	0.771

TABLE X
AV CLASSIFICATION AND DISAMBIGUATION ACCURACY OF TECHNIQUES DEVELOPED FOR FUNDUS CAMERA IMAGES TESTED ON UWFOV SLO IMAGES.

	[21]	[18]
ZONE C	0.774	0.759
WIDE - zone C	0.721	0.734
WIDE - central	0.708	0.697
WIDE - peripehery	0.643	0.625
TASCFORCE - zone C	0.762	0.748
TASCFORCE - central	0.723	0.699
TASCFORCE - peripehery	0.664	0.648
WIDE 3-degree points	0.769	0.738
WIDE 4-degree points	0.805	0.749
TASCFORCE 3-degree points	0.772	0.744
TASCFORCE 4-degree points	0.801	0.753

- ENet = elastic net [38].
- LASSO = least absolute shrinkage and selection operator [39].
- OLS = ordinary least squares.
- RLS = regularized least squares.
- LR = linear regression.
- LSVM = linear support vector machine.
- RBF = RBF-kernel support vector machine.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of all co-workers and volunteers involved in the Tayside Screening for the Prevention of Cardiac Events (TASCFORCE) study, including Professor Jill Belch and Dr Roberta Littleford, University of Dundee, Dr Charles Weller and Mrs Patricia Martin, NHS Tayside. Thanks to Matteo Barbieri and Dr Annalisa Barla, University of Genoa. Thanks to Estrada and colleagues for making their dataset publicly available. This study was funded by a studentship from Optos plc (Dunfermline, UK).

REFERENCES

[1] M. Abramoff, M. Garvin, and M. Sonka, "Retinal imaging and image analysis," *Biomedical Engineering, IEEE Reviews in*, vol. 3, pp. 169–208, 2010.

[2] T. MacGillivray, E. Trucco, J. Cameron, B. Dhillon, J. Houston, and E. Van Beek, "Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions," *The British journal of radiology*, vol. 87, no. 1040, p. 20130832, 2014.

[3] S. McGrory, J. R. Cameron, E. Pellegrini, C. Warren, F. N. Doubal, I. J. Deary, B. Dhillon, J. M. Wardlaw, E. Trucco, and T. J. MacGillivray, "The application of retinal fundus camera imaging in dementia: A systematic review," *Alzheimerś & Dementia: Diagnosis, Assessment & Disease Monitoring*, pp. –, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.dadm.2016.11.001

[4] M. K. Ikram, C. Y. Cheung, M. Lorenzi, R. Klein, T. L. Z. Jones, and T. Y. W. and, "Retinal vascular caliber as a biomarker for diabetes microvascular complications," *Diabetes Care*, vol. 36, no. 3, pp. 750–759, feb 2013.

[5] S. Frost, Y. Kanagasingam, H. Sohrabi, J. Vignarajan, P. Bourgeat, O. Salvado, V. Villemagne, C. C. Rowe, S. L. Macaulay, C. Szoeke, K. A. Ellis, D. Ames, C. L. Masters, S. Rainey-Smith, and R. N. Martins, "Retinal vascular biomarkers for early detection and monitoring of alzheimerś disease," *Translational Psychiatry*, vol. 3, no. 2, p. e233, feb 2013.

[6] T. Y. Wong, R. Klein, A. R. Sharrett, B. B. Duncan, D. J. Couper, B. E. Klein, L. D. Hubbard, and F. J. Nieto, "Retinal arteriolar diameter and risk for hypertension," *Ann. Intern. Med.*, vol. 140, no. 4, pp. 248–255, Feb 2004.

[7] M. K. Ikram, J. C. Witteman, J. R. Vingerling, M. M. Breteler, A. Hofman, and P. T. de Jong, "Retinal vessel diameters and risk of hypertension: The Rotterdam study," *Hypertension*, vol. 47, no. 2, pp. 189–194, 2006.

[8] T. T. Nguyen and T. Y. Wong, "Retinal vascular manifestations of metabolic disorders," *Trends in Endocrinology & Metabolism*, vol. 17, no. 7, pp. 262–268, sep 2006.

[9] R. Kawasaki, J. Xie, N. Cheung, E. Lamoureux, R. Klein, B. E. Klein, M. F. Cotch, A. R. Sharrett, S. Shea, and T. Y. Wong, "Retinal microvascular signs and risk of stroke: the Multi-Ethnic Study of Atherosclerosis (MESA)," *Stroke*, vol. 43, no. 12, pp. 3245–3251, Dec 2012.

[10] "Fundus photographic risk factors for progression of diabetic retinopathy: Etdrs report number 12," *Ophthalmology*, vol. 98, no. 5, pp. 823–833, may 1991.

[11] R. H. Webb and G. W. Hughes, "Scanning laser ophthalmoscope," *Biomedical Engineering, IEEE Transactions on*, no. 7, pp. 488–492, 1981.

[12] J. Sherman, G. Karamchandani, W. Jones, S. Nath, and L. Yannuzzi, *Panoramic Ophthalmoscopy: Optomap Images and Interpretation.* New Jersey: Slack Incorporated, 2007.

[13] R. Estrada, M. Allingham, P. Mettu, S. Cousins, C. Tomasi, and S. Farsiu, "Retinal artery-vein classification via topology estimation," *Medical Imaging, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[14] V. S. Joshi, J. M. Reinhardt, M. K. Garvin, and M. D. Abramoff, "Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks," *PLoS ONE*, vol. 9, no. 2, p. e88061, 2014.

[15] E. Grisan and A. Ruggeri, "A divide et impera strategy for automatic classification of retinal vessels into arteries and veins," in *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, vol. 1, Sept 2003, pp. 890–893 Vol.1.

[16] S. Vázquez, N. Barreira, M. G. Penedo, M. Ortega, and A. Pose-Reino, "Improvements in retinal vessel clustering techniques: towards the automatic computation of the arterio venous ratio," *Computing*, vol. 90, no. 3-4, pp. 197–217, 2010.

[17] M. Saez, S. González-Vázquez, M. González-Penedo, M. A. Barceló, M. Pena-Seijo, G. C. de Tuero, and A. Pose-Reino, "Development of an automated system to classify retinal vessels into arteries and veins," *Computer methods and programs in biomedicine*, vol. 108, no. 1, pp. 367–376, 2012.

[18] D. Relan, T. MacGillivray, L. Ballerini, and E. Trucco, "Automatic retinal vessel classification using a least square-support vector machine in vampire," in *36th Annual International Conference of the IEEE EMBS Engineering in Medicine and Biology Society (EMBC), Chicago, USA*, 2014, pp. 142–145.

[19] C. Muramatsu, Y. Hatanaka, T. Iwase, T. Hara, and H. Fujita, "Automated selection of major arteries and veins for measurement of arteriolar-to-venular diameter ratio on retinal fundus images," *Computerized Medical Imaging and Graphics*, vol. 35, no. 6, pp. 472–480, 2011.

[20] M. Niemeijer, X. Xu, A. V. Dumitrescu, P. Gupta, B. Van Ginneken, J. C. Folk, and M. D. Abramoff, "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 11, pp. 1941–1950, 2011.

[21] A. Zamperini, A. Giachetti, E. Trucco, and K. S. Chin, "Effective features for artery-vein classification in digital fundus images," in *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, June 2012, pp. 1–6.

[22] Q. Mirsharif, F. Tajeripour, and H. Pourreza, "Automated characterization of blood vessels as arteries and veins in retinal images," *Computerized Medical Imaging and Graphics*, vol. 37, no. 7, pp. 607–617, 2013.

[23] J. Gama, "A linear-bayes classifier," in *Advances in Artificial Intelligence, International Joint Conference, 7th Ibero-American Conference on AI, 15th Brazilian Symposium on AI, IBERAMIA-SBIA 2000, Atibaia, SP, Brazil, November 19-22, 2000, Proceedings*, 2000, pp. 269–279.

[24] B. Al-Diri, A. Hunter, D. Steel, and M. Habib, "Automated analysis of retinal vascular network connectivity," *Computerized Medical Imaging and Graphics*, vol. 34, no. 6, pp. 462–470, 2010.

[25] Y. Huang, J. Zhang, and Y. Huang, "An automated computational framework for retinal vascular network labeling and branching order analysis," *Microvascular research*, vol. 84, no. 2, pp. 169–177, 2012.

[26] K. Rothaus, X. Jiang, and P. Rhiem, "Separation of the retinal vascular graph in arteries and veins based upon structural knowledge," *Image and Vision Computing*, vol. 27, no. 7, pp. 864–875, 2009.

[27] Q. P. Lau, M. L. Lee, W. Hsu, and T. Y. Wong, "Simultaneously identifying all true vessels from segmented retinal images," *IEEE Trans Biomed Eng*, vol. 60, no. 7, pp. 1851–1858, Jul 2013.

[28] B. Dashtbozorg, A. M. Mendonca, and A. Campilho, "An automatic graph-based approach for artery/vein classification in retinal images," *IEEE Trans Image Process*, vol. 23, no. 3, pp. 1073–1083, Mar 2014.

[29] F. Rotaru, S.-I. Bejinariu, R. Luca, and C. Niţă, "Retinal vessel labeling method," in *E-Health and Bioengineering Conference (EHB), 2015*. IEEE, 2015, pp. 1–4.

[30] Q. Hu, M. K. Garvin, and M. D. Abramoff, "Automated construction of arterial and venous trees on retinal images using topological and intensity information," *Investigative Ophthalmology & Visual Science*, vol. 56, no. 7, pp. 5264–5264, 2015.

[31] K. Eppenhof, E. Bekkers, T. T. Berendschot, J. P. Pluim, and B. M. ter Haar Romeny, "Retinal artery/vein classification via graph cut optimization," in *Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop*. University of Iowa, sep 2015. [Online]. Available: https://doi.org/10.17077/omia.1035

[32] R. Estrada, C. Tomasi, S. Schmidler, and S. Farsiu, "Tree topology estimation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[33] D. E. Croft, J. van Hemert, C. C. Wykoff, D. Clifton, M. Verhoek, A. Fleming, and D. M. Brown, "Precise montaging and metric quantification of retinal surface area from ultra-widefield fundus photography and fluorescein angiography," *Ophthalmic Surg Lasers Imaging Retina*, vol. 45, no. 4, pp. 312–317, 2014.

[34] http://medicine.dundee.ac.uk/tascforce.

[35] E. Pellegrini, G. Robertson, E. Trucco, T. J. MacGillivray, C. Lupascu, J. van Hemert, M. C. Williams, D. E. Newby, E. J. van Beek, and G. Houston, "Blood vessel segmentation and width estimation in ultra-wide field scanning laser ophthalmoscopy," *Biomed Opt Express*, vol. 5, no. 12, pp. 4329–4337, Dec 2014.

[36] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Commun. ACM*, vol. 27, no. 3, pp. 236–239, Mar. 1984.

[37] A. Cavinato, L. Ballerini, E. Trucco, and E. Grisan, "Spline-based refinement of vessel contours in fundus retinal images for width estimation," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, April 2013, pp. 872–875.

[38] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[39] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267–288, 1994.

[40] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1124–1137, September 2004.

[41] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary amp; region segmentation of objects in n-d images," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, 2001, pp. 105–112 vol.1.

[42] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, June 1947.

[43] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[44] P. Samer and S. Urrutia, "A branch and cut algorithm for minimum spanning trees under conflict constraints," *Optimization Letters*, vol. 9, no. 1, pp. 41–55, 2015.

[45] A. Darmann, U. Pferschy, J. Schauer, and G. J. Woeginger, "Paths, trees and matchings under disjunctive constraints," *Discrete Applied Mathematics*, vol. 159, no. 16, pp. 1726 – 1735, 2011, 8th Cologne/Twente Workshop on Graphs and Combinatorial Optimization (CTW 2009).